Title	Dependencies in a Partially Specified Relation (データ・セマンティクスの理論と実際に関する研究)
Author(s)	TANAKA, Yuzuru
Citation	数理解析研究所講究録 (1982), 461: 236-264
Issue Date	1982-06
URL	http://hdl.handle.net/2433/103129
Right	
Type	Departmental Bulletin Paper
Textversion	publisher

数理解析研究所講究録
461巻 1982年 236-264

Title : Dependencies in a Partially Specified Relation

Author : Yuzuru TANAKA
Faculty of Engineering,
Hokkaido University,
Sapporo, 060 JAPAN

Abstract :

The current schema design theories assume that a database may be regarded as a single relation. Usually, this is hardly acceptable. However, if a relation is allowed to have unspecified items, any database can be represented as such a partial relation. Since the normalization theory is for relations, it needs to be extended so that it may become applicable to partial relations. As extensions of dependencies, natural dependencies are defined. They are axiomatized, and the set of axioms is proved its completeness. The schema design based on natural dependencies solves not only the problems caused by a universal relation assumption but also the update anomalies caused by decomposition of a database.

1. Introduction.

While the theories on database schema design that are based on dependencies among attributes of a relation have been much criticized their inapplicability to the actual design of practical databases, the growing dimension and complexity of databases are increasing the needs for a CAD system for the design of database schemata. In order to automatize a major portion of design processes, such a CAD system needs as its basis a mathematical design algorithm based on a complete axiomatic system. The purpose of this paper is to fill up an alleged gap between theory and practice of the schema design.

The current design theories are based on the-so-called normalization theory, which was originally applied only to a single relation. Normalization was first proposed by E. F. Codd [CODD72]. It decomposes a relation to decrease update anomalies. The decomposition increases the locality of update operations, and hence it saves trouble in the execution of update requests. Unfortunately, the same theory has been applied to a database, which is not always a single relation but a set of relations. The schema design theories originated from a hasty conclusion that normalization is also applicable to databases. Therefore, they have to assume that an object database can be regarded as a single relation. This assumption is called a universal relation assumption. Obviously, it is hardly acceptable from a practical point of view. The alleged gap between theory and practice originates in this assumption.

This paper will extend the definition of a relation in Section 3 to allow some of its tuples to have attributes whose values are not specified. Such an extended relation is called a partial relation. Instead of regarding a database as a relation, we will regard it as a partial relation. This assumption imposes no essential restrictions on object databases. The definition of dependencies is also extended in Section 4 to describe dependency structures in partial relations. New dependencies are called natural dependencies. The replacement of dependencies by their corresponding natural dependencies will make the normalization theory and the design algorithms also applicable to partial relations. Section 5 will axiomatize natural dependencies, and prove its completeness. Section 7 will show this extention will solve various problems caused by update operations.

2. Problems in the Conventional Normalization Theory.

Schema design theories based on the conventional normalization theory assume the universal relation assumption. The examples in this section will show the unreality of this assumption.

Let R(A, B, C) as shown in Fig. 1 (a) be a relation satisfying a functional dependency $B \rightarrow C$. In this section, R is taken as an example database that can be regarded as a single relation. Schema design theories decompose the database R to yield as its schema two projections of R, i.e., R_1 and R_2 in Fig. 1 (b). There are two different major approaches to the design of schemata. One is the decomposition approach, and the other is the synthesis approach. In this case, however, the result is independent from which of the two approaches is applied.

Problems will arise when we try to update R(A, B, C) that is actually stored as a set of two relations R_1 and R_2. The following two update requests will explain the problems;

(1) Delete the relationship that B is 'c' and C is 'e'.

(2) Delete the value 'c' from the values of B.

The execution of (1) will change R_1 and R_2 as in Fig. 1 (c), while the execution of (2) will change them as in (f). The result (c) has no corresponding universal relation over {A, B, C}. A table in (d) represents the result (c), however, it is not a relation since it has an unspecified item. Such a table with unspecified items is called a partial relation. Since unspecified items may be regarded to take a special value '⊥', a table (d) may be identified with (e). This example indicates the possibilities that even a very simple update request may make a database to lose its universal relation even if it initially satisfies the universal relation assumption.

There is another noticeable point in (d). Although the projection of a tuple (b, c, e) to the attribute set {A, B} is preserved in the result of the update, its projection to {A, C} disappears from the result. This seems to reflect a tacit understanding that, if $B \rightarrow C$ holds, the value of C is not specified with the corresponding value of B remaining unspecified.

The execution of (2) will change two relations in (b) to two tables in (f), which cause another problem. Although the universal table for this result should be the table (g), another table (h) is also decomposed to (f). Such ambiguity is caused by an update operation that yield a partial relation with unspecified items in its key attributes. In the relational model of databases, partial relations of this kind have been tacitly prohibited.

The example update (1) indicates that universal relations should be allowed to have unspecified items. The second observation on (1) and the example (2) indicate that dependencies used for decomposition of a partial relation should not be specified independently from the appearance of unspecified items in this partial relation. Dependencies and the appearance of unspecified items seem to be closely related.

The E-R model that is comparatively accepted by designers of databases descriminates between two types of attributes, i.e., entities and properties. It considers only those dependencies $X \to Y$ and $X \to\to Y$ whose determinant X has no properties. The values of properties can not be specified independently without specifying the corresponding entities or the relationship among entities. Therefore, if Y has no entities, dependencies $X \to Y$ or $X \to\to Y$ satisfies the condition:

> "In each tuple, its X values are completely specified if there exists some attribute A in Y whose value is specified."

It is proved in this paper that decomposition of partial relations by such dependencies satisfying the above condition yields no such undesirable partial relations that have unspecified items in their key attributes. Therefore, it seems desirable to define dependencies in partial relations to satisfy the above condition.

3. Partial Relations

Let $[X \to Y]$ denote a set of all the total functions from a set X to another set Y. A set of all the partial functions from X to Y is denoted by $[X \to Y]'$. Let Ω be a finite set and D an enumerable set. A relation over (Ω, D) with Ω as its attribute set and D as its value set might be interpreted as a subset of $[\Omega \to D]$. As an extension of this interpretation, a partial relation over (Ω, D) is defined as a subset of $[\Omega \to D]'$. In distinction from partial relations, ordinary relations are said to be total. An element of $[\Omega \to D]$ is called a (total) tuple over (Ω, D), while an element of $[\Omega \to D]'$ is called a partial tuple. Let '$\perp$' be a special value called a bottom that is not in D, and $\underline{D}$ a union of D and $\{\perp\}$. For each partial tuple μ in $[\Omega \to D]'$, a total tuple $\underline{\mu}$ in $[\Omega \to \underline{D}]$ is defined as

$$\underline{\mu}(x) = \begin{cases} \mu(x) & \text{if } \mu(x) \neq \text{undefined}, \\ \perp & \text{if } x \in \Omega \text{ and } \mu(x) = \text{undefined}, \\ \text{undefined} & \text{otherwise.} \end{cases}$$

The support of a tuple $\underline{\mu}$ over $(\Omega, \underline{D})$ that is denoted by $s(\underline{\mu})$ is defined as

$$s(\underline{\mu}) = \{A \mid A \in \Omega \wedge \underline{\mu}(A) \neq \perp\}.$$

A tuple $\underline{\mu}$ is said to be superior to $\underline{\mu}'$ if

$$x \in \Omega \ (\underline{\mu}'(x) \neq \perp) \supset (\underline{\mu}(x) = \underline{\mu}'(x)),$$

which is denoted by $\underline{\mu} > \underline{\mu}'$. The difference $\underline{\mu} - \underline{\mu}'$ of two tuples $\underline{\mu}$ and $\underline{\mu}'$ in $[\Omega \to \underline{D}]$ is also a tuple in $[\Omega \to \underline{D}]$ defined as

$$(\underline{\mu} - \underline{\mu}')(x) = \begin{cases} \underline{\mu}(x) & \text{if } \underline{\mu} \not> \underline{\mu}' \\ \underline{\mu}(x) & \text{if } \underline{\mu} > \underline{\mu}' \text{ and } x \notin s(\underline{\mu}') \\ \perp & \text{otherwise.} \end{cases}$$

For each partial relation r over (Ω, D), a corresponding total relation $\underline{r}$ over $(\Omega, \underline{D})$ is defined as

$$\underline{r} = \{\underline{\mu} \mid \mu \in r\}.$$

By $\omega(r)$ is referred to the attribute set of $\underline{r}$. Since a relation $\underline{r}$ may be identified with r, $\underline{r}$ is also called a partial relation. Similarly, a tuple $\underline{\mu}$ is also called a partial tuple.

The restriction of a partial tuple $\underline{\mu}$ over $(\Omega, \underline{D})$ within a subset X of Ω is an element of $[X \to \underline{D}]$ defined as

$$\underline{\mu}|_X(x) = \begin{cases} \underline{\mu}(x) & \text{if } x \in X, \\ \text{undefined} & \text{otherwise.} \end{cases}$$

The projection of a partial relation $\underline{r}$ with respect to an attribute set X is a suset of $[\Omega \cap X \to D]'$ defined as

$$[X]\underline{r} = \begin{cases} \{\underline{\mu}|_X \mid \underline{\mu} \in \underline{r}\} & \text{if } X \subset \Omega, \\ \phi & \text{otherwise.} \end{cases}$$

We define a directed join of two relations $\underline{r}$ and $\underline{s}$ as

$$\underline{r} \triangleright \underline{s} = \{\underline{\mu} \mid (\underline{\mu} \in [\omega(r) \cup \omega(s) \to \underline{D}]) \wedge (\underline{\mu}|_{\omega(r)} \in \underline{r}) \wedge (\underline{\mu}|_{\omega(s)} \in \underline{s}) \wedge \forall_{A \in \omega(s)-\omega(r)} \forall_{B \in \omega(s) \cap \omega(r)} ((\underline{\mu}(A) \neq \perp) \supset (\underline{\mu}(B) \neq \perp))\}.$$

The join of two tatal relations R and S is denoted R*S.

4. Natural Dependencies

As the dependencies are defined with respect to total relations, we will define natural dependencies with respect to partial relations.

Def. 4.1.

Let r, X and Y be a partial relation and two subsets of $\omega(r)$. An existential dependency (ED) $X \overset{e}{\rightarrow} Y$ is defined as

$$\underline{r} \text{ sat } X \overset{e}{\rightarrow} Y \quad \text{iff} \quad \forall B \in Y \ \exists A \in X \ \forall \underline{\mu} \in \underline{r} \quad (\underline{\mu}(A) \neq \perp) \supset (\underline{\mu}(B) \neq \perp),$$

where $\underline{r}$ sat C denotes that $\underline{r}$ satisfies the condition C.

Def. 4.2.

Let r, X and Y be the same as in Def. 4.1. A natural functional dependency (nFD) $X \Rightarrow Y$ is defined as

$$r \text{ sat } X \Rightarrow Y \quad \text{iff} \quad (\underline{r} \text{ sat } X \rightarrow Y) \wedge (\forall A \in Y \ \ \underline{r} \text{ sat } \{A\} \overset{e}{\rightarrow} X).$$

Def. 4.3.

Let r, X and Y be the same as in Def. 4.1. A natural multivalued dependency (nMVD) $X \Rightarrow\!\!\!\!\Rightarrow Y$ is defined as

$$r \text{ sat } X \Rightarrow\!\!\!\!\Rightarrow Y \quad \text{iff} \quad (\underline{r} \text{ sat } X \twoheadrightarrow Y) \wedge (\forall A \in Y \ \ \underline{r} \text{ sat } \{A\} \overset{e}{\rightarrow} X).$$

A natural dependency $X \Rightarrow Y$ ($X \Rightarrow\!\!\!\!\Rightarrow Y$) is a dependency $X \rightarrow Y$ ($X \twoheadrightarrow Y$) with the implications that the value of any attribute in Y is not specified with the corresponding value of some attribute in X remaining unspecified.

As a relation R satisfies that

$(R \text{ sat } X \rightarrow\rightarrow Y)$ iff $R = [XY]R*[X(\Omega-Y)]R$,

a partial relation satisfies a similar relation as shown in the next theorem.

Th. 4.1.

Let r, X and Y be the same as in Def. 4.1. A partial relation is decomposable if it satisfies a nontrivial natural dependency, i.e.,

$(r \text{ sat } X \twoheadrightarrow Y)$ iff $\underline{r} = [X(\Omega-Y)]\underline{r} \triangleright [XY]\underline{r}$.

(proof) Obvious from the definitions.

If total relations are concerned, natural dependencies might be identified with the corresponding dependencies.

5. Axiomatization of Natural Dependencies

FD's and MVD's are known to satisfy the following set of axioms.

FD1 (Reflexivity)	if $Y \subset X$ then $X \to Y$.
FD2 (Augmentation)	if $Z \subset W$ and $X \to Y$ then $XW \to YZ$.
FD3 (Transitivity)	if $X \to Y$ and $Y \to Z$ then $X \to Z$.
MVD0 (Complementation)	if $X \to\to Y$ then $X \to\to \Omega - Y$.
MVD1 (Augmentation)	if $Z \subset W$ and $X \to\to Y$ then $XW \to\to YZ$.
MVD2 (Transitivity)	if $X \to\to Y$ and $Y \to\to Z$ then $X \to\to Z - Y$.
FD-MVD1	if $X \to Y$ then $X \to\to Y$.
FD-MVD2	if $X \to\to Y$ and $(\Omega - Y) \to Y$ then $X \to Y$.

Lemma 5.1.
The above set of axioms is complete with respect to FD's and MVD's.
(proof) See [BEER77].

Now, we axiomatize the properties of ED's.

ED1. (Reflexivity)	if $Y \subset X$ then $X \overset{e}{\to} Y$.
ED2. (Augmentation)	if $Z \subset W$ and $X \overset{e}{\to} Y$ then $XW \overset{e}{\to} YZ$.
ED3. (Transitivity)	if $X \overset{e}{\to} Y$ and $Y \overset{e}{\to} Z$ then $X \overset{e}{\to} Z$.

The above set of axioms for ED's is essentially the same as the axioms for ED's.

Th. 5.2.

ED's satisfy the above axioms ED1 ~ ED3.

(proof) Trivial.

Th. 5.3.

The above set of axioms for ED's is complete with respect to ED's.

(proof)

Let Λ be an arbitrary set of ED's among subsets of Ω. A set of all the ED's inferable from Λ using ED1 ~ ED3 is called the closure of Λ. It is denoted by $\Lambda^{\dagger}$. The set of axioms is complete if, for any Λ, there always exist a value set $\underline{D}$ and a partial relation $\underline{r}$ over $(\Omega, \underline{D})$ such that $\underline{r}$ satisfies all the ED's in $\Lambda^{\dagger}$ but not any ED's other than those in $\Lambda^{\dagger}$. Here we show how such $\underline{D}$ and $\underline{r}$ can be constructed for an arbitrarily given Λ. Let Ω be $\{A_1, A_2, \ldots, A_n\}$. For each $X \subset \Omega$, a set X^* is defined as

$$X^* = \{B \mid X \overset{e}{\rightarrow} B \in \Lambda^{\dagger}\},$$

and a value set D with 2^n distinct elements is defined as

$$D = \{a_X \mid X \subset \Omega\}.$$

For each $X \subset \Omega$, a tuple $\underline{\mu}_X$ over $(\Omega, \underline{D})$ is defined as

$$\underline{\mu}_X(A) = \begin{cases} a_X & \text{if } A \in X^*, \\ \perp & \text{otherwise.} \end{cases}$$

A partial relation $\underline{r}$ is constructed as

$$\underline{r} = \{\underline{\mu}_X \mid X \subset \Omega\}.$$

Obviously, $\underline{r}$ satisfies $\Lambda^{\dagger}$. Suppose that $\underline{r}$ satisfies an ED $f : X \overset{e}{\rightarrow} Y$ that is not in $\Lambda^{\dagger}$. Since an ED $X \overset{e}{\rightarrow} X^*$ is included in $\Lambda^{\dagger}$, a set $Y-X^*$ must not be empty. Let B be an element of

$Y-X^*$. Since $\underline{r}$ satisfies $X \overset{e}{\to} Y$, it also satisfies $X \overset{e}{\to} B$. This implies that

$$\exists A \in X \ \forall \underline{\mu} \in \underline{r} \quad (\underline{\mu}(A) \neq \bot) \supset (\underline{\mu}(B) \neq \bot),$$

which is equivalent to the condition;

$$\exists A \in X \ \forall \underline{\mu} \in \underline{r} \quad (\underline{\mu}(B) = \bot) \supset (\underline{\mu}(A) = \bot). \tag{5.1}$$

While $B \notin X^*$ implies $\underline{\mu}_X(B) = \bot$, the definition of $\underline{\mu}_X$ implies $\underline{\mu}_X(A) = a_X$. This contradicts the condition (5.1). Therefore, $\underline{r}$ does not satisfy f. (Q.E.D.)

Natural dependencies are axiomatized by adding the following two axioms to the axioms FD1 ~ FD3, MVD0 ~ MVD2, FD-MVD1, FD-MVD2, and ED1 ~ ED3.

nFD	$X \Rightarrow Y$	iff	$(X \to Y) \wedge (\forall A \in Y \quad A \overset{e}{\to} X)$
nMVD	$X \Rightarrow\!\!\!\Rightarrow Y$	iff	$(X \twoheadrightarrow Y) \wedge (\forall A \in Y \quad A \overset{e}{\to} X)$

Th. 5.4.

A set of axioms consisting of ED1 ~ ED3, MVD0 ~ MVD2, FD-MVD1, FD-MVD2, ED1 ~ ED3, nFD, and nMVD is complete with respect to nFD's and nMVD's.

(proof)

Let Γ be a set of nFD's and nMVD's among subsets of Ω. We define two sets as

$$\Gamma_0 = \{X \to Y \mid X \Rightarrow Y \in \Gamma\} \cup \{X \twoheadrightarrow Y \mid X \Rightarrow\!\!\!\Rightarrow Y \in \Gamma\}.$$

$$\Gamma_1 = \{A \overset{e}{\to} X \mid (X \Rightarrow Y \in \Gamma \vee X \Rightarrow\!\!\!\Rightarrow Y \in \Gamma) \wedge (A \in Y)\}.$$

A set of ED's inferable from Γ_1 by ED1 ~ ED3 is denoted by $\Gamma_1^{\dagger}$, and a set of FD's and MVD's inferable from Γ_0 by the remaining axioms is denoted by $\Gamma_0^{\dagger}$.

The above set of axioms is complete if, for any Γ and Ω, there exists a value set $\underline{D}$ and a partial relation $\underline{r}$ over $(\Omega, \underline{D})$ such that $\underline{r}$ satisfies $\Gamma^{\dagger}$ but not any natural dependencies other than those in $\Gamma^{\dagger}$. Here we show how to construct such a value set D and an example partial relation $\underline{r}$ for an arbitrarily given Ω and Γ. From Lemma 5.1, there exist, for each Ω and Γ, a value set D_0 and an example relation R_0 over (Ω, D_0) such that R_0 satisfies ${\Gamma_0}^{\dagger}$ but not any dependencies other than those in ${\Gamma_0}^{\dagger}$. From Th. 5.3, there exist a value set D_1 and an example partial relation $\underline{r}_1$ that satisfies ${\Gamma_1}^{\dagger}$ but not any ED's other than those in ${\Gamma_1}^{\dagger}$. Since the elements of D_1 defined in Th. 5.3 are independent from D_0, D_0 and D_1 can be made mutually disjoint. Let $\underline{D}$ and $\underline{r}$ over $(\Omega, \underline{D})$ be defined as

$$\underline{D} = D_0 \cup \underline{D}_1,$$

$$\underline{r} = R_0 \cup \underline{r}_1.$$

Then $\underline{r}$ satisfies $\Gamma^{\dagger}$ but not any natural dependencies other than those in $\Gamma^{\dagger}$. This is proved as follows.

Let f be a natural dependency $X \Rightarrow Y$ (or $X \twoheadrightarrow\!\!\!\Rightarrow Y$) that is arbitrarily chosen from $\Gamma^{\dagger}$. Since the corresponding dependency holds, ${\Gamma_0}^{\dagger}$ includes $X \rightarrow Y$ (or $X \twoheadrightarrow Y$). Therefore, the definition of R_0 implies that

$$R_0 \text{ sat } X \rightarrow Y \text{ (or } X \twoheadrightarrow Y \text{).} \qquad (5.2)$$

On the other hand, $f \in \Gamma^{\dagger}$ implies that

$$\forall A \in Y \quad A \overset{e}{\rightarrow} X \in {\Gamma_1}^{\dagger},$$

which further implies that

$$\forall A \in Y \ \forall Z \subset \Omega \quad (A \in Z) \supset (X \subset Z^*)$$

because A* includes X and Z* includes A*.

Therefore, for any $Z \subset \Omega$, if $\underline{\mu}_Z$ in $\underline{r}_1$ which was defined in the proof of Th. 5.3 satisfies $\underline{\mu}_Z(A) \neq \perp$, then, for any B in X, $\underline{\mu}_Z(B)$ is equal to $\underline{\mu}_Z(A)$. Since a_Z does not appear in any tuples in $\underline{r}_1$ other than $\underline{\mu}_Z$, this implies that

$$\forall A \in Y \quad \underline{r}_1 \text{ sat } X \rightarrow A,$$

or equivalently

$$\underline{r}_1 \text{ sat } X \rightarrow Y. \tag{5.3.a}$$

Obviously, this also implies

$$\underline{r}_1 \text{ sat } X \rightarrow\!\!\rightarrow Y. \tag{5.3.b}$$

Since the value sets of R_0 and $\underline{r}_1$ are mutually disjoint, conditions (5.2) and (5.3) imply that

$$\underline{r} \text{ sat } X \rightarrow Y \text{ (or } X \rightarrow\!\!\rightarrow Y \text{)}. \tag{5.4}$$

Now we will prove under the same assumption that

$$\forall A \in Y \quad \underline{r} \text{ sat } A \overset{e}{\rightarrow} X.$$

Since $X \Rightarrow Y$ (or $X \Rightarrow\!\!\!\Rightarrow Y$) is an element of $\Gamma^{\dagger}$, $\Gamma_1{}^{\dagger}$ includes $A \overset{e}{\rightarrow} X$ for any A in Y. Therefore, the definition of $\underline{r}_1$ implies

$$\forall A \in Y \quad \underline{r}_1 \text{ sat } A \overset{e}{\rightarrow} X. \tag{5.5}$$

Since R_0 is a total relation, it is obvious that

$$\forall A \in Y \quad R_0 \text{ sat } A \overset{e}{\rightarrow} X. \tag{5.6}$$

The conditions (5.5) and (5.6) imply that

$$\forall A \in Y \quad \underline{r} \text{ sat } A \overset{e}{\rightarrow} X. \tag{5.7}$$

From (5.4) and (5.7), it can be concluded that

$$r \text{ sat } X \Rightarrow Y \text{ (or } X \Rightarrow\!\!\!\Rightarrow Y \text{)}. \tag{5.8}$$

The remaining part of our proof is to show that $\underline{r}$ does not satisfy any natural dependency f that is not in $\Gamma^{\dagger}$. Let f be $X \Rightarrow Y$ (or $X \Rightarrow\!\!\!\Rightarrow Y$) that is not in $\Gamma^{\dagger}$. This assumption implies either

$$X \rightarrow Y \quad \text{(or } X \rightarrow\!\!\rightarrow Y \text{)} \notin \Gamma_0{}^{\dagger}, \quad \text{or}$$

$\exists A \varepsilon Y \quad A \overset{e}{\to} X \not\in \Gamma_1^{\dagger}$.

If $X \to Y$ (or $X \twoheadrightarrow Y$) is not included in $\Gamma_0^{\dagger}$, then the definition of R_0 implies that

$$\neg(R_0 \text{ sat } X \to Y \text{ (or } X \twoheadrightarrow Y \text{))}, \qquad (5.9)$$

and hence it is proved from the disjointness of D_0 and $\underline{D}_1$ that

$$\neg(r \text{ sat } X \Rightarrow Y \text{ (or } X \Rightarrow\!\!\Rightarrow Y \text{))}.$$

Otherwise, there exists A in Y such that $A \overset{e}{\to} X$ is not in $\Gamma_1^{\dagger}$. Therefore, the definition of $\underline{r}_1$ implies that

$$\neg(\underline{r}_1 \text{ sat } A \overset{e}{\to} X), \qquad (5.10)$$

and hence it is proved that

$$\neg(r \text{ sat } X \Rightarrow Y \text{ (or } X \Rightarrow\!\!\Rightarrow Y \text{))}.$$

Thus the set of axioms as shown above is complete with respect to natural dependencies. (Q.E.D.)

6. The Computation of the Natural Dependency Closures.

The proof of Th. 5.4 gives a hint how to compute $\Gamma^\dagger$ from an arbitrarily given Γ. The following is an algorithm for this computation.

ALGORITHM

step 1. Obtain Γ_0 and Γ_1 from Γ.

step 2. Compute ${\Gamma_0}^\dagger$ from Γ_0 using axioms for FD's and MVD's.

step 3. Compute ${\Gamma_1}^\dagger$ from Γ_1 using axioms for ED's.

step 4. Compute $\Gamma^\dagger$ as

$$\Gamma^\dagger = \{X \rightarrow \{A\} \mid (X \rightarrow \{A\} \in {\Gamma_0}^\dagger) \wedge (A \not\supseteq X \in {\Gamma_1}^\dagger)\}$$
$$\cup \{X \twoheadrightarrow Y \mid (X \twoheadrightarrow Y \in {\Gamma_0}^\dagger) \wedge (\forall A \in Y \quad A \not\supseteq X \in {\Gamma_1}^\dagger)\}.$$

The algorithms to compute ${\Gamma_0}^\dagger$ have already been studied by several authors. Here we show an algorithm using the canonical representation of dependencies [TANA79]. For a partition $\{X, Y_0, Y_1, \ldots, Y_n\}$ of Ω, a representation

$$X: [Y_0] Y_1 | Y_2 | \ldots | Y_n$$

is called a canonical representation. It denotes a set of dependencies

$X \rightarrow Y_0$, and

$\forall i \quad X \twoheadrightarrow Y_i$.

Let C be a set of all the canonical representations over Ω. The dependency base of Γ_0 that is denoted by $\tilde{\Gamma}_0$ is a subset of C that is defined as

$$\tilde{\Gamma}_0 = \{\delta \mid (\delta \in C) \wedge (\Gamma_0 \vdash \delta)$$
$$\wedge \neg(\exists \delta' \in C - \{\delta\} \quad (\Gamma_0 \vdash \delta') \wedge (\delta' \vdash \delta))\}.$$

Since, for each element f in $\Gamma_0^{\dagger}$, there always exist an element δ in $\tilde{\Gamma}_0$ and an element f' in δ such that $f' \vdash f$, the computation of $\Gamma_0^{\dagger}$ might be replaced by that of $\tilde{\Gamma}_0$. The dependency base $\tilde{\Gamma}_0$ can be computed by using the following rules.

rule 1.

if $\Gamma_0 \vdash X:[Y_0]Y_1|Y_2|\ldots|Y_n$ and $\Gamma_0 \vdash U:[V_0]V_1|V_2|\ldots|V_m$

then $\Gamma_0 \vdash W:[Z_0]Z_1|Z_2|\ldots|Z_m|Z_{m+1}$, where

$W = X \cup (Y_i \cap U)$,

$Z_0 = Y_0 \cup (Y_i \cap V_0) - W$,

$Z_j = Y_i \cap V_j$ for $1 \leq j \leq m$,

$Z_{m+1} = \Omega\text{-}W\text{-} \bigcup_{0 \leq j \leq m} Z_j$.

rule 2.

$U:[V_0]V_1|V_2|\ldots|V_m \vdash X:[Y_0]Y_1|Y_2|\ldots|Y_n$

iff $(U \subset X) \wedge (V_0 \supset Y_0)$

$\wedge (\forall i \;\; \exists I \subset \{0, 1, 2, .., m\} \;\; X \cup Y_i = U \cup (\bigcup_{j \in I} V_j))$.

If MVD's are not concerned, each canonical representation $X:[Y_0]Y_1$ might be replaced with $X:[Y_0]$. The computation of $\Gamma_1^{\dagger}$ in step 3 is essentially the same as the computation of the closure of functional dependencies. Therefore, its dependency base $\tilde{\Gamma}_1$ is defined. The Computation of $\Gamma_1^{\dagger}$ might be replaced by that of $\tilde{\Gamma}_1$.

An example is given below to show how to apply this algorithm.

Example 6.1.

The computation of $\Gamma^{\dagger}$ for such Ω and Γ given as

$$\Omega = \{A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q\},$$

$$\Gamma = \{G \Rightarrow DK,\ AC \Rightarrow OP,\ H \Rightarrow AB,\ AB \twoheadrightarrow CDEFGKLM,\ C \twoheadrightarrow DEFGKN,$$
$$D \twoheadrightarrow AHKLN,\ F \twoheadrightarrow ABG,\ I \twoheadrightarrow JQ\}$$

is shown below.

step 1.

$$\Gamma_0 = \{G \to DK,\ AC \to OP,\ H \to AB,\ AB \to\to CDEFGKLM,\ C \to\to DEFGKN,$$
$$D \to\to AHKLN,\ F \to\to ABG,\ I \to\to JQ\},$$

$$\Gamma_1 = \{D \overset{e}{\to} G,\ K \overset{e}{\to} G,\ O \overset{e}{\to} AC,\ P \overset{e}{\to} AC,\ A \overset{e}{\to} H,\ B \overset{e}{\to} H,$$
$$C \overset{e}{\to} AB,\ D \overset{e}{\to} AB,\ E \overset{e}{\to} AB,\ F \overset{e}{\to} AB,\ G \overset{e}{\to} AB,$$
$$K \overset{e}{\to} AB,\ L \overset{e}{\to} AB,\ M \overset{e}{\to} AB,\ D \overset{e}{\to} C,\ E \overset{e}{\to} C,\ F \overset{e}{\to} C,$$
$$G \overset{e}{\to} C,\ K \overset{e}{\to} C,\ N \overset{e}{\to} C,\ A \overset{e}{\to} D,\ H \overset{e}{\to} D,\ K \overset{e}{\to} D,$$
$$L \overset{e}{\to} D,\ N \overset{e}{\to} D,\ A \overset{e}{\to} F,\ B \overset{e}{\to} F,\ G \overset{e}{\to} F,\ J \overset{e}{\to} I,\ Q \overset{e}{\to} I\}.$$

step 2.

$\tilde{\Gamma}_0$: G:[ABDKOP]H|L|N|IJQ|CEFM

H:[ABOP]N|IJQ|CDEFGKLM

AB:[OP]H|N|IJQ|CDEFGKLM

C:[ABOP]H|N|IJQ|DEFGK|L|M

D:[ABKOP]H|L|N|IJQ|CEFGM

F:[ABDKOP]G|H|L|N|IJQ|CEM

I:JQ|ABCDEFGHKLMNOP

step 3.

$\tilde{\Gamma}_1$: A:[BCDFGH], B:[ACDFGH], C:[ABDFGH], D:[ABCFGH],

F:[ABCDGH], G:[ABCDFH], H:[ABCDFG],

E:[ABCDFGH], J:[I], K:[ABCDFGH], L:[ABCDFGH],

M:[ABCDFGH], N:[ABCDFGH], O:[ABCDFGH],

P:[ABCDFGH], Q:[I]

<u>step 4.</u>

$\Gamma^{\dagger}$: $G \Rightarrow ABDKOP$, $G \twoheadrightarrow H$, $G \twoheadrightarrow L$, $G \twoheadrightarrow N$, $G \twoheadrightarrow CEFM$,

$H \Rightarrow ABOP$, $H \twoheadrightarrow N$, $H \twoheadrightarrow CDEFGKLM$,

$AB \Rightarrow OP$, $AB \twoheadrightarrow H$, $AB \twoheadrightarrow N$, $AB \twoheadrightarrow CDEFGKLM$,

$C \Rightarrow ABOP$, $C \twoheadrightarrow H$, $C \twoheadrightarrow N$, $C \twoheadrightarrow DEFGK$, $C \twoheadrightarrow L$, $C \twoheadrightarrow M$,

$D \Rightarrow ABKOP$, $D \twoheadrightarrow H$, $D \twoheadrightarrow L$, $D \twoheadrightarrow N$, $D \twoheadrightarrow CEFGM$,

$F \Rightarrow ABDKOP$, $F \twoheadrightarrow G$, $F \twoheadrightarrow H$, $F \twoheadrightarrow L$, $F \twoheadrightarrow N$, $F \twoheadrightarrow CEM$,

$I \twoheadrightarrow JQ$.

It should be noticed there exist some dependencies that are not natural dependencies. An MVD $G \twoheadrightarrow IJQ$ is such an example.

7. Updates of a Partial Relation

The insertion of a partial tuple μ over (Ω, D) to a partial relation r over (Ω, D) is a simple process to change $\underline{r}$ to $\underline{r} \cup \{\underline{\mu}\}$. However, the deletion of a partial tuple ν over (Ω, D) from r is not so simple. In this paper, it is defined as a process to change $\underline{r}$ to

$$\{\underline{\nu}'-\underline{\nu} \mid \underline{\nu}' \in \underline{r}\}.$$

The difference of two tuples is already defined in Section 3. An example of such deletion is shown in Fig. 2.

Assume that Ω is an attribute set satisfying natural dependencies Γ. For a subset X of Ω, X* and *X are defined as

$$X^* = \{A \mid (A \in \Omega) \wedge (\Gamma \vdash X \overset{e}{\rightarrow} A)\}, \text{ and}$$

$$^*X = \{A \mid (A \in \Omega) \wedge (\{A\}^* \cap X \neq \phi)\}.$$

These two sets X* and *X are respectively called an insertion base and a deletion base of X. Suppose that a partial tuple μ is to be inserted to a partial relation over (Ω, D) satisfying Γ. A partial tuple μ must satisfy existential dependencies in Γ_1. Therefore, if its value is specified at an attribute A, it is also specified at any attributes in $\{A\}^*$. A subset X of Ω is said to be insertion compatible if X* is equal to X. If μ is to be inserted, its support $s(\mu)$ should be insertion compatible.

Suppose that a partial tuple ν is to be deleted from a partial relation r over (Ω, D) satisfying Γ. Let $\underline{\nu}'$ be a tuple in $\underline{r}$ satisfying $\underline{\nu}' > \underline{\nu}$. Suppose that there exists an attribute A in *X such that $\underline{\nu}'(A)$ will not become '$\perp$' after the deletion. Since $A \overset{e}{\rightarrow} \{A\}^*$ holds in Γ_1, there exists some attribute B in $\{A\}^* \cap X$ and $\underline{\nu}'(B)$ will not be '$\perp$'. However, since ν is to be deleted, $\underline{\nu}'(B)$ should become '$\perp$'. Therefore, the value of ν' should become unspecified at all the attributes in $*s(\nu)$. A deletion base of $s(\nu)$ is a maximum set of attributes to which the deletion of a partial tuple ν might propagate.

Th. 7.1

For any subset X of Ω and any insertion compatible set Y, $(Y-{*X})$ is also insertion compatible, i.e.,

$$(Y-{*X})^* = Y-{*X}.$$

(proof)

If there exists A in $(Y-{*X})^*-(Y-{*X})$, then A is in $Y \cap {*X}$ because the insertion compatibility of Y implies that

$$(Y-{*X})^*-(Y-{*X}) \subset Y-(Y-{*X}) = Y \cap {*X}.$$

Therefore, A satisfies

$$Y-{*X} \overset{e}{\rightarrow} A, \quad \text{and} \quad \exists\, B \in X \quad A \overset{e}{\rightarrow} B.$$

Since each ED in Γ_1 has only one attribute in its determinant, there exists some C in $Y-{*X}$ such that $C \overset{e}{\rightarrow} A$. This implies $C \overset{e}{\rightarrow} B$, and further implies C is in $*X$. This contradicts the condition that $C \in Y-{*X}$. Therefoer, $(Y-{*X})$ should be insertion compatible.. (Q.E.D.)

Def. 7.1

Let Ω_1 and Ω_2 be two subsets of Ω. A pair (Ω_1, Ω_2) is a decomposition of Ω if it satisfies

$$\Gamma \vdash (\Omega_1 \cap \Omega_2) \twoheadrightarrow \Omega_1-\Omega_2 \quad \text{or} \quad \Gamma \vdash (\Omega_1 \cap \Omega_2) \twoheadrightarrow \Omega_2-\Omega_1.$$

Th. 7.2.

Let (Ω_1, Ω_2) be a decomposition of Ω. If X is insertion compatible and neither of Ω_1 nor Ω_2 includes X, then the set of join attributes is included in X, i.e.,

$$\Omega_1 \cap \Omega_2 \subset X.$$

(proof)

If neither of $X \subset \Omega_1$ nor $X \subset \Omega_2$ holds, then neither of the sets defined as

$$X_1 = X \cap (\Omega_1 - \Omega_2),$$
$$X_2 = X \cap (\Omega_2 - \Omega_1)$$

is empty. Let X_3 be $X-(X_1 \cup X_2)$. Since (Ω_1, Ω_2) be a decomposition, either $\Omega_1 \cap \Omega_2 \twoheadrightarrow \Omega_1 - \Omega_2$ or $\Omega_1 \cap \Omega_2 \twoheadrightarrow \Omega_2 - \Omega_1$ holds. If $\Omega_1 \cap \Omega_2 \twoheadrightarrow \Omega_1 - \Omega_2$ holds, then $X_1{}^*$ includes $\Omega_1 \cap \Omega_2$. Otherwise, $X_2{}^*$ includes $\Omega_1 \cap \Omega_2$. Therefore, if X^* is X, then X includes $\Omega_1 \cap \Omega_2$. (Q.E.D.)

Suppose that a relation $\underline{r}$ is decomposable by an ordinary dependency and that it is actually stored as a set of its two projections $[\Omega_1]\underline{r}$ and $[\Omega_2]\underline{r}$. The insertion of a partial tuple $\underline{\mu}$ which crosses these two relations will cause a problem if $\underline{\mu}$ is not specified its value at some attribute in $\Omega_1 \cap \Omega_2$. In such a case, if $\underline{\mu}|_{\Omega_1}$ and $\underline{\mu}|_{\Omega_2}$ are separately inserted to $[\Omega_1]\underline{r}$ and $[\Omega_2]\underline{r}$, the join of these two relations can not reconstruct $\underline{\mu}$ since $\underline{\mu}|_{\Omega_1}$ and $\underline{\mu}|_{\Omega_2}$ lack the values of the join attributes. However, if the natural dependencies in $\underline{r}$ are well specified and $\underline{r}$ is decomposed by one of them, such a problem will not occur. In such a situation, $s(\mu)$

must be insertion compatible and hence, from Th. 7.2, it must include $\Omega_1 \cap \Omega_2$. Therefore, the directed join can reconstruct $\underline{\mu}$ from $\underline{\mu}|_{\Omega_1}$ and $\underline{\mu}|_{\Omega_2}$.

As it was shown in Section 2, the above situation also causes a problem in deletion operations. If a relation $\underline{r}$ is inappropriately decomposed, some kind of deletion operation yields an undesirable tuple in $\underline{r}$ whose value is not specified at some of the join attributes. However, if the decomposition is performed by a natural dependency, it is guaranteed by Th. 7.1 that such a situation will not occur.

This desirable property of the decompositions by natural dependencies is preserved for further decompositions. Therefore, if a schema is designed by decomposing a universal partial relation and each decomposition uses a natural dependency, then no updates will make any constituent relation in this schema to have unspecified items in its join attributes.

8. Conclusions

The current schema design theories are based on the-so-called normalization theory, which was originally applied to a single relation and has no theoretical foundation for its applicability to a partial relation. Thus the design theories had to assume that an object database may be regarded as a single relation. Usually, this assumption is hardly acceptable. However, any database can be regarded as a single partial relation. Essentially, the normalization theory is based on two concepts, a dependency among attributes and decomposition of a relation. As their natural extensions, similar concepts for partial relations has been defined in this paper. The extensions are natural since the new concepts degenerate into the original concepts if a partial relation happens to be a relation. The naturalness of these extensions makes it possible to use the essential part of the conventional theories with a little modification.

The join operation that is used as a basis of decomposition has been replaced by the directed join operation. An extended dependency is called a natural dependency. A natural dependency $X \Rightarrow Y$ (or $X \Rrightarrow Y$) is basically a dependency $X \rightarrow Y$ (or $X \twoheadrightarrow Y$) with implications that, if some partial tuple is specified its value at some attribute in Y, it must be specified its value at all the attributes in X. Section 5 has axiomatized natural dependencies and proved the completeness of the set of axioms. Section 6 has shown an algorithm to compute the closure of an arbitrarily given set of natural dependencies.

Updates of a database generally causes various problems. If a database is decomposed and a partial tuple is not specified its value at some of the join attributes, the insertion of this tuple across several relations is impossible. The similar troubles also occur in deletion operations. However, it hs been proved that such a trouble will never occur if natural dependencies are well specified with respect to an object database and decomposition is done by one of them. Therefore, the schema design using natural dependencies solves not only the problem caused by a universal relation assumption but also the update anomalies caused by decomposition of a database.

References

[BEER77] Beeri, C., R. Fagin, and J. H. Howard, "A complete axiomatization for functional and multivalued dependencies," ACM SIGMOD International Symposium on Management of Data, 1977, pp.47-61.

[CODD72] Codd, E. F., "Further normalization of the data base relational model," in Data Base Systems (R. Rustin, ed.), Prentice Hall, Englewood Cliffs, N.J., 1972, pp.33-64.

[TANA79] Tanaka, Y., "Logical design of a relational schema and integrity of a data base," in Data Base Architecture (Bracchi and Nijssen, ed.), North-Holland, Amsterdam, 1979, pp.297-316.

(a)

R	A	B	C
	a	b	d
	b	c	e

$B \rightarrow C$

(b)

R_1	A	B
	a	b
	b	c

R_2	B	C
	b	d
	c	e

(1) Delete the relationship that B is 'c' and C is 'e'.

(2) Delete the value 'c' from the values of B.

(c)

R_1	A	B
	a	b
	b	c

R_2	B	C
	b	d

(f)

R_1	A	B
	a	b
	b	$\perp$

R_2	B	C
	b	d
	$\perp$	e

Universal Relation ?!

(d)

R	A	B	C
	a	b	d
	b	c	

(g)

R	A	B	C
	a	b	d
	b	$\perp$	e

(h)

R	A	B	C
	a	b	d
	b	$\perp$	$\perp$
	$\perp$	$\perp$	e

(e)

R	A	B	C
	a	b	d
	b	c	$\perp$

Fig. 1 Problems caused by the execution of updates on a designed database.

$$\Gamma = \phi$$

R	A	B	C	D
	e	f	g	h
	i	j	⊥	⊥
	⊥	a	b	⊥
	e	a	⊥	c
	⊥	a	b	c
	d	a	b	c

$\Downarrow$

deletion of (⊥ a b c)

$\Downarrow$

R	A	B	C	D
	e	f	g	h
	i	j	⊥	⊥
	⊥	a	b	⊥
	e	a	⊥	c
	d	⊥	⊥	⊥

Fig. 2 Deletion of a partial tuple from a partial relation.